Modeling Frequency of Truck Crashes on Limited-Access Highways

by Niranga Amarasingha and Sunanda Dissanayake

Freight can be efficiently transported between most locations in the U.S. using large trucks. Involvement of large trucks in crashes can cause much damage and serious injuries, due to their large sizes and heavy weights. The purpose of this study was to identify the relationships between large truck crashes and traffic and geometric characteristics on limited access highways. Crash and traffic and geometric-related data for Kansas were utilized to develop a Poisson regression model and a negative binomial regression model for understanding the relationships. Based on model-fitting statistics, the negative binomial model was found to be the better model, which was used to identify the important characteristics. By addressing identified factors, safety could be promoted through introduction of appropriate engineering improvements.

INTRODUCTION

In the United States, large trucks provide a convenient mode for the movement of freight from origin to destination. The American Trucking Association reported a 47% increase in registered large trucks and a 65% increase in their miles traveled from 1988 to 2008 (ATA 2012). In 2009, large trucks accounted for 4% of all registered vehicles and 10% of total vehicle miles traveled in the U.S. (NHTSA 2009). Trucks with gross vehicle weight greater than 10,000 pounds are typically considered large trucks, and 296,000 of such trucks were involved in traffic crashes on U.S. roadways during 2009 (NHTSA 2009). There were 3,380 fatalities and 74,000 injuries reported due to those large truck crashes that year. Also, according to 2009 statistics, large trucks represented about 7% of vehicles in fatal crashes, 2% of vehicles in injury crashes, and 3% of vehicles in property-damage-only crashes (NHTSA 2009). Involvement of large trucks in crashes can cause much damage and serious injuries due to their large sizes and heavy weights. In 2009, occupants of large trucks comprised only 22% of fatalities resulting from fatal large truck crashes, while 78% of the fatalities occurred outside the truck to pedestrians, cyclists, and, primarily, occupants of passenger vehicles (NHTSA 2009).

Injuries and severity of injuries that occur in a crash increase exponentially with vehicle speed (Stuster 1999). However, long distance freight transportation requires large trucks to have access to interstate and state highways and operate at higher speeds. Also, drivers may face vehicle control challenges or difficulties while driving large trucks on interstate or state highways at high speeds. Interstates and urban highways serve a diverse combination of passenger vehicle traffic, local delivery truck traffic, and long-haul truck traffic.

Analysis of large truck crash data indicates there are traffic and highway geometric characteristics associated with large truck crashes (Miaou 1994). Highway geometric design features such as a horizontal curvature, vertical grade, lane width, lane type, shoulder width, shoulder type, and median are engineering factors which might be used to reduce the number of large truck crashes. One of many important aspects of highway safety research is developing crash prediction models to quantify the relationship between traffic and geometric characteristics and the number of crashes observed. Identifying the effects of traffic and geometric characteristics is important to promote safety by introducing engineering improvements. The focus of this research was to understand and evaluate the effects of both traffic conditions and site characteristics on the occurrence of large truck crashes.

LITERATURE REVIEW

Several previous studies have investigated the relationship between crash rates and traffic and geometric design features. A number of crash frequency models have been developed for large truck crashes exclusively. Mohamedshah et al. (1993) investigated traffic and geometric-related variables that affect truck crashes using data from the Highway Safety Information System (HSIS). Multivariate logistic models for truck crashes on interstates and two-lane rural roads were developed considering truck crash data in Utah from 1980 to 1989. As indicated by the statistically significant variables in the interstate model, truck crashes were primarily affected by horizontal curvature and vertical gradient. Vertical gradient is inclination of road surface to the horizontal plane while horizontal curvature can be defined as a measure of the sharpness of a horizontal curve. When values of horizontal curvature or vertical grade increases, the number of truck crashes increases. For two-lane rural roads, as indicated by the statistically significant variables in the model, truck crashes were affected by shoulder width and horizontal curvature. With the increase of the horizontal curvature, the number of truck crashes increases; however, shoulder width and number of truck crashes have a negative relationship.

Miaou (1994) evaluated the relationship between truck crashes and geometric design features of road sections using Poisson regression, Zero-Inflated Poisson regression, and negative binomial regression models. Data were obtained from the HSIS, which included 1,643 large truck crashes occurring on Utah highway sections within the five-year period from 1985 to 1989. Estimated regression parameters from all three models were quite consistent in terms of estimated relative frequencies of truck crashes across road sections. The developed models were then evaluated based on estimated regression parameters, overall goodness-of-fit, predicted relative frequency of truck crashes, sensitivity to the inclusion of short road sections, and estimated total number of truck crashes. Evaluation results showed that Poisson regression models were best to use as the initial model for developing the relationship, while other forms of models could be explored if the over-dispersion (i.e., the variance of crash frequency in the dataset exceeds the mean of the crash frequency) of crash data is found in the Poisson model. According to estimated coefficients of the significant variables, truck crashes increase with the increase of the annual average daily traffic (AADT) per lane, horizontal curvature, and vertical grade while number of truck crashes decrease with the increase of percentage of trucks in the traffic.

Schneider et al. (2009) developed a negative binomial regression model using crash data from Ohio to investigate the effect of rural two-lane horizontal curves on truck crashes at non-intersection locations. Data were obtained from the Ohio Department of Public Safety and Ohio Department of Transportation's roadway inventory files, which includes all heavy-duty truck crashes related to single- and multi-vehicle crashes on horizontal curves. This study further investigated implementation of Bayesian methods on model performance. Impact of shoulder width, curve radius, curve length, and traffic parameters on truck crashes were considered in the model development. The significant variables in the final model were length of horizontal curve, truck annual daily traffic (ADT), passenger ADT, and degree of horizontal curve. Each of these variables had a positive relationship with the number of truck crashes. The developed model was used to target improvements to specific roadways. The model could also be used to identify truck crashes that may occur in the future due to volume increases. The authors pointed out the need for improved models to accommodate other, non-volume-related contributing factors to truck crashes to improve the truck-crash-frequency prediction.

Virginia crash data were used by Joshua and Garber (1990) to find the quantitative relationship between traffic and geometric variables, and the probability of occurrence of large truck crashes.

Geometric data such as number of lanes, lane and shoulder widths, and vertical and horizontal alignments were collected directly from the sites at which a large number of truck-related crashes occurred. Multiple linear and Poisson regression analyses were carried out in order to predict the number of truck crashes, where the Poisson regression model was found to be capable of better describing the relationship. It indicated that the rate of change of slope (change in vertical grade divided by the length of the highway segment), average daily traffic, percent of trucks, and speed differential between trucks and non-trucks had significantly influenced the number of truck crashes. Increase of each of these variables indicated more truck crashes.

Daniel et al. (2002) developed a crash prediction model for truck crashes on route sections with signalized intersections. Crash data were obtained from New Jersey accident records, and volume and geometric data were obtained by reviewing straight-line diagrams and contract drawings of the roadway. A Poisson regression model and a negative binomial regression model were developed. Coefficients of the negative binomial model were comparable with those for the Poisson regression model with some exceptions. Coefficients of both models showed significant impact based on segment length, AADT, length of vertical grade, number of lanes, number of signals within the segment, and pavement width on truck crash frequency on selected roadways. According to both models, with the increase of segment length, AADT, number of lanes, and number of signals within the segment, the number of truck crashes increase. The increase in the length of vertical grade and pavement width showed decreased number of truck crashes.

DATA

Crash data from 2005 to 2010 were obtained from the Kansas Department of Transportation (KDOT), which were utilized for analysis in this study. These data, included in the Kansas Accident Reporting System (KARS) database, comprise all police-reported crashes in Kansas. For this study, large truck crash records on limited-access highways were extracted by making the query from all crashes from 2005–2010 for the state of Kansas. From 2005 to 2010, 5,392 large trucks were involved in crashes on limited-access highways. After identifying these large truck crashes, information to locate each crash on the highway was obtained from the Control Section Analysis System (CANSYS) database.

The CANSYS database, maintained by KDOT, is a highway inventory system that includes many traffic- and geometric-related details of national and state highways in Kansas. Data from 2005 to 2010 were obtained for limited-access highways, and sections were defined based on homogeneity of road segments and data availability. The selected sections were homogenous in terms of number of lanes, horizontal curvature, median width, AADT, truck AADT percent, lane width, shoulder width, and existence of rumble strips. Additionally, variables such as functional class, section length, and year were considered in the analysis. For this study, data on vertical grade (i.e., inclination of road surface to the horizontal plane) were provided by KDOT from construction drawings, as vertical grades are not frequently updated in the CANSYS database.

A total of 16,853 roadway segments were initially identified where the length varied from 0.10 miles to 19.87 miles, with an average segment length of 0.81 miles. Data were reviewed and sections which had speed limits lower than 55 mph and lengths shorter than 0.25 miles were discarded. A total of 7,273 roadway segments were considered for further analysis. Table 1 shows summary characteristics of road segments used in the analysis. All roadway sections were divided-roadway sections, as the focus of this study is on limited-access roads.

Table 1: Traffic- and Geometric-Related Characteristics of Limited-Access Highway Sections

Variable	Description	Sections No.	Sections %	Variable	Description	Sections No.	Sections %
Section Length (SL, in miles)	0.25 ≤ *SL* < 0.50	2,466	33.91	*Right Rumble Strips*	Yes	4,492	61.76
	0.50 ≤ *SL* < 1.00	2,053	28.23		No	2,781	38.24
	1.00 ≤ *SL* < 2.00	1,302	17.90	*Inside Rumble Strips*	Yes	4,200	57.75
	2.00 ≤ *SL* < 3.00	522	7.18		No	3,073	42.25
	3.00 ≤ *SL* < 4.00	265	3.64	*Right Shoulder Width (ft)*	0	68	0.93
	4.00≤ *SL*	665	9.14		2	24	0.33
Speed Limit (mph)	55	122	1.68		6	46	0.63
	60	769	10.57		8	195	2.68
	65	1,486	20.43		9	8	0.11
	70	4,896	67.32		10	6,838	94.02
Median Width (MW, in ft)	*MW*<10	92	1.26		12	94	1.29
	10 ≤ *MW* < 20	2,087	28.70	*Inside Shoulder Width (ft)*	0	2,845	39.12
	20 ≤ *MW* < 30	349	4.80		2	41	0.56
	30 ≤ *MW* < 40	3,997	54.96		3	8	0.11
	40≤ *MW*	748	10.28		4	32	0.44
Functional Class	Expressways	1,211	16.65		6	2,970	40.84
	Rural interstate	3,351	46.07		7	16	0.22
	Urban interstate	2,711	37.27		8	126	1.73
AADT per lane (veh/day/ lane)	Less than 1,000	302	4.15		9	511	7.03
	1,000 - 2,000	2,965	40.77		10	696	9.57
	2,000 - 3,000	1,224	16.83		12	28	0.38
	3,000 - 4,000	629	8.65	*Horizontal Curve*	Curve	660	9.07
	4,000 - 5,000	437	6.01		Straight	6,613	90.93
	More than 5,000	1,716	23.59	*Vertical Grade*	Level	6,795	93.43
AADT of Large Truck Count (veh/day)	Less than 1,000	1,549	21.30		Grade	478	6.57
	1,000 - 2,000	4,728	65.01	*Number of Lanes*	4	5,855	80.50
	2,000 - 3,000	768	10.56		6	1,220	16.77
	More than 3,000	228	3.13		8	198	2.72

Shoulder widths and rumble strips were recorded separately for inside (or left) and outside (or right) lanes in a given direction. Absolute values of horizontal curvature and vertical grade on each homogeneous section were used for the modeling. Sections having either positive or negative horizontal curvature or vertical grade are more dangerous for trucks than other sections where the roadway is level and straight. For example, since trucks may not be able to maintain normal, prevailing traffic speeds on steep upgrades, they may lead to sudden braking by the following vehicles, resulting in overturning or rear-end crashes. Total number of crashes occurring within each segment was determined by combining crash data and CANSYS data. About 35% of the road segments had at least one large truck crash, regardless of truck configurations and crash-severity type, while the remaining segments did not have any reported large truck crashes during the years that were being considered. Large crashes are potentially affected by human factors as well, but data related to human factors were unavailable or not possible to be aggregated based on individual road sections. Similar situation exists for contributory causes for each section as well. The omitted and unavailable factors were kept consistent to better understand the effect of traffic and geometric relationships that are being investigated.

METHODOLOGIES

Various statistical models could be considered for identifying relationships between number of large truck crashes and geometric and traffic characteristics. Because of the random and discrete nature of crashes, Poisson regression has long been considered as a good starting point for frequency modeling (Miaou 1994).

Poisson Regression Model

Poisson regression model is appropriate for dependent variables that have non-negative integer values such as 0, 1, 2… Hence, in most cases, count data could be precisely analyzed by Poisson regression (Pedan 2001). More details of Poisson regression analysis can be found in Long (1997).

The Poisson regression model was proposed by Miaou (1994) to find the relationship between vehicle crashes and geometric design features of road sections, such as lane width, shoulder width, horizontal curvature, and lane width. The Poisson regression model proposed by Miaou (1994) is given by:

$$(1)\quad P(Y_i = y_i) = p(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad (i = 1,2,3,\dots n;\ y_i = 0,1,2,3,\dots)$$

where,

- i = a roadway segment. The same roadway segments in different sample periods are considered as separate roadway segments.
- y_i = the number of large truck crashes for a year for roadway segment i.
- $P(y_i)$ = probability of the occurrence of y_i large truck crashes for a year on roadway segment i.
- μ_i = mean value of large truck crashes occurring for a year as:

$$(2)\quad \mu_i = E(Y_i) = \vartheta_i[e^{\sum_{j=1}^{k} x_{ij}\beta_j}]$$

where,

- x_{ij} = the j^{th} independent variable for roadway segment i,
- β_j = the coefficient for the j^{th} independent variable, and
- ϑ_i = traffic exposure for roadway segment i.

Associated with each roadway segment i, x_i independent variables describe geometric characteristics, traffic conditions, and other relevant attributes. Traffic exposure, which is the amount of large truck travel during the sample year, can be computed as:

$$(3)\quad \vartheta_i = 365 \times AADT_i \times T\%_i \times l_i$$

where,

- ϑ_i = traffic exposure on segment i,
- $AADT_i$ = Annual Average Daily Traffic (vehicles/day),
- $T\%_i$ = percentage of large trucks in traffic stream, and
- l_i = length of the road segment.

This model assumes the number of large truck crashes for a given time period for roadway segments (Y_i; i = 1,2, ... n) are independent of each other and Poisson distributed with mean. The expected number of large truck crashes $E(Y_i)$ is proportional to large truck travel ϑ_i. The model ensures that the crash frequency is positive, using an exponential function given by:

$$(4)\quad \lambda_i = \frac{E(Y_i)}{\vartheta_i} = \exp\left(x_i' \beta\right)$$

where,

λ_i = exposure-based number of large truck crashes,
$E(Y_i)$ = expected number of large truck crashes,
x_i' = transpose of covariate vector x_i , and
β = vector of unknown regression parameters.

One important property of Poisson regression is that it restricts the mean and variance of the distribution to be equal. This can be written as:

$$(5)\quad \mathrm{Var}(y_i) = E(y_i) = \mu_i$$

where,

μ_i = mean of the response variable y_i ,
$E(y_i)$ = expected number of response variable y_i , and
$\mathrm{Var}(y_i)$ = variance of response variable y_i .

If this equality does not hold, the data are said to be either underdispersed or overdispersed, and the resulting parameter estimates will be biased. If the overdispersion is not captured in the analysis, the standard errors are underestimated and, hence, it becomes an overstatement of significance in hypothesis testing (Pedan 2001). If the model fits the data, both deviance and Pearson Chi-Square statistics divided by the degrees of freedom are approximately equal to one. The deviance is the likelihood-ratio statistic for comparing the model to the saturated model, which explains all the variation in the data. Values greater than one indicate the variance is an overdispersion, while values smaller than one indicate an underdispersion. It is possible to account for overdispersion with respect to the Poisson model by introducing a scale (dispersion) parameter into the relationship between the variance and the mean (Pedan 2001).

Another way to address overdispersion, if it exists, is the consideration of a distribution that permits more flexible modeling of the variance. The negative binomial regression model is more appropriate for overdispersed data because it relaxes the constraints of equal mean and variance.

Negative Binomial Regression Model

The following details of negative binomial regression models related to highway large truck crashes were described in many studies (Miaou 1994, Schneider et al. 2009, and Daniel et al. 2002). Consider a set of n highway sections of a limited-access highway. Let Y_i be a random variable representing the number of large trucks involved in crashes on highway section i during the analysis period. Further, assume the amount of large truck travel or large truck exposure on this highway section, V_i , is also a random variable estimated through a highway sampling system. Associated with each highway section i is a $k \times 1$ vector of explanatory variables, denoted by $\underline{x}_i = (x_{i1} = 1, x_{i2}, \ldots \ldots x_{ik})'$, describing its geometric characteristics, traffic conditions, and other relevant attributes. Given V_i , and x_i , large truck crash involvements Y_i, $i = 1,2,3, \ldots \ldots, n$, are postulated to be independent, and each is Poisson distributed as follows (Miaou 1994):

(6) $P(Y_i = y_i) = \dfrac{(\lambda_i \vartheta_i)^{y_i} e^{-\lambda_i \vartheta_i}}{y_i!}$

where,
λ_i = large truck crash involvement and
ϑ_i= exp (random error).

If a loglinear rate function is used as follows, the model becomes the negative binomial regression model that gives the relationship between the expected number of crashes occurring at the i^{th} section with K number of parameters.

(7) $\lambda_i = \exp(\beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{iK} + \varepsilon_i)$

where,

λ_i	= number of large truck crashes on limited-access highway section i, with negative binomial distribution conditional on ε_i,
β_0	= constant term,
$\beta_1, \ldots \ldots, \beta_n$	= estimated parameters in vector form,
$X_1, \ldots \ldots, X_n$	= explanatory variables in vector form, and
$\in_i$	= random error; exp(($\in_i$) is distributed as gamma with mean 1 and variance α^2.

In the case of the Poisson regression model, coefficients β_i are estimated by maximizing the log likelihood $\log_e L(\beta)$.

Assessment of the Models

In order to assess the adequacy of models, the basic descriptive statistics for the event count data first need to be investigated (Pedan 2001). The models developed using the relevant statistically significant variables are further tested for goodness-of-fit, which includes deviance statistics and Pearson Chi-Square statistics.

Deviance statistics are used to assess the fit of the model and overdispersion. These statistics are sometimes referred to as the likelihood ratio statistics or G-squared value. The G-squared value is the sum of deviance, and is defined as the change in deviance between the fitted model and the model with a constant term and no covariates. The G-squared statistics are given by (Agresti 2007):

(8) $G^2 = 2\Sigma_{=1}^{n} y_i \ln(y_i/E(y_i))$

where,

G^2	= deviance,
y_i	= observed number of large truck crashes,
E(y_i)	= expected number of large truck crashes, and
n	= number of road sections.

If this statistic is significant, then the covariates contribute significantly to the model. If not, other covariates and/or other error distributions need to be considered. Deviance is approximately a chi-squared random variable with degrees of freedom (DF) equal to the number of observations (n) minus the number of parameters (p). A value of the deviance over ($n - p$) that is degrees of freedom, suggests the model is overdispersed due to missing variables and/or a non-Poisson form. Thus, when deviance divided by degrees of freedom is significantly larger than one, overdispersion is indicated.

Pearson Chi-Squared statistics are used to assess the presence of overdispersion in the model and are given in equation (9) (Agresti 2007):

$$(9)\quad x^2 = \Sigma_{i=1}^{n} \frac{(y_i - \lambda_i)^2}{\lambda_i}$$

where,

y_i = observed number of large truck crashes
λ_i = expected number of large truck crashes, and
n = number of road sections.

If value of the Chi-Squared statistics over degrees of freedom is larger than 1, overdispersion is also indicated. If Pearson Chi-Square statistics divided by degrees of freedom and deviance statistics divided by degrees of freedom are both closer to one, it indicates a better model fit.

ANALYSIS RESULTS

With consideration given to variables used in the literature and data availability, candidate variables were selected and the definitions of variables considered for individual road sections, along with the descriptive statistics, are presented in Table 2. A total of 17 explanatory variables were selected to be considered in the model. The existence of right rumble strip and inside (left) rumble strip were considered as categorical variables. As the number of lanes varies from section to section, AADT per lane was considered in the modeling. Maximum horizontal curvature was 4% per 100 ft of arc (degrees of curvature), while the maximum grade was 3.35%. There was considerable variation of risk across the years due to long-term trends, and changes in omitted variables such as road surface conditions and weather. Therefore, year-to-year changes in overall large truck crashes were captured using yearly dummy variables in the model.

Poisson Regression Model

A Poisson regression model was developed, taking into account the previously explained variables. Goodness-of-fit statistics showed that deviance/DF and Pearson Chi-Square statistic/DF were both slightly higher than 1.00, which suggested more variability among counts than would be expected for Poisson distribution. The descriptive data also indicated the overdispersion of data showed the mean number of crashes in a section was 0.66 with a variance of 1.34, as given in Table 2.

One of the most common reasons for data being overdispersed is that μ_i parameters vary not only with measured covariates, but with latent and uncontrolled factors. Hence, without any adjustment for overdispersion, the Poisson model was not quite adequate to describe the occurrence of large truck crashes on limited-access highways in Kansas. Accordingly, the Poisson model was adjusted for overdispersion by including a scale (dispersion) parameter, as presented in Table 3. The scale parameter was estimated by considering a ratio of the Pearson Chi-Square to its associated degrees of freedom. The estimated scale parameter was 1.2145 and scaled Pearson Chi-Square was fixed to one.

In Table 3, the coefficient of each independent variable influencing the large truck crashes in the model gave the size of the exponential effect of a particular variable on the number of large truck crashes. The coefficients of continuous variables bearing a positive sign indicated an increase in large truck crashes with an increase of the variable, while a negative sign indicated a decrease in large truck crashes with an increase of the variable. The coefficient of dummy or indicator variables bearing a positive sign indicates the dummy or indicator variable switch from 0 to 1; that is an increase of crashes. A unit change in the variable would affect large truck crashes by an exponential power of that variable coefficient, if all other variables were kept constant. The variables that were

Table 2: Variable Definitions for Limited-Access Highway Large Truck Frequency Modeling

Variable	Description	Notation and Definition	Min	Max	Mean	Std. Dev.
TRUCK_NUM	Number of large truck crashes in a section	y_i	0	32	0.66	1.34
SEC_LEN	Section length (miles)	l_i	0.25	19.87	1.51	2.20
TR_EXPO	Truck miles or truck exposure (log scale)	$\vartheta_i = (365 \times AADT_i \times T\%_i \times l_i) / 1000$	6.10	14.04	10.63	1.12
Y_2005	Dummy variable for year 2005	x_{i1} = 1, if road section is in year 2005; 0 otherwise	0	1	0.15	0.35
Y_2006	Dummy variable for year 2006	x_{i2} = 1, if road section is in year 2006; 0 otherwise	0	1	0.15	0.36
Y_2007	Dummy variable for year 2007	x_{i3} = 1, if road section is in year 2007; 0 otherwise	0	1	0.15	0.36
Y_2008	Dummy variable for year 2008	x_{i4} = 1, if road section is in year 2008; 0 otherwise	0	1	0.16	0.36
Y_2009	Dummy variable for year 2009	x_{i5} = 1, if road section is in year 2009; 0 otherwise	0	1	0.18	0.38
Y_2010	Dummy variable for year 2010	x_{i6} = 1, if road section is in year 2010;0 otherwise	0	1	0.21	0.40
AADT	Surrogate variable to indicate traffic conditions or traffic density	$x_{i7} = (AADT_i / no.\ of\ lanes_i) / 1000$	0.15	13.83	3.35	2.48
HC	Horizontal curvature (absolute value in degrees per 100ft arc)	x_{i8}	0	4	0.11	0.43
VG	Vertical grade (absolute value in percent)	x_{i10}	0	3.35	0.23	0.58
IN_SHOULD	Inside shoulder width	x_{i12}	0	12	9.83	1.17
R_SHOULD	Right-side shoulder width	x_{i13}	0	12	6.99	1.88
TRUCK	Percent large trucks in the traffic stream (log scale)	x_{i14}	1.42	39.16	10.63	1.12
IN_RUMBLE	Dummy variable for inside rumble strip	x_{i5} = 1, if inside rumble strip exists; 0 otherwise	0	1	0.61	0.48
R_RUMBLE	Dummy variable for right rumble strip	x_{i5} = 1, if right rumble strip exists; 0 otherwise	0	1	0.57	0.49
SPEED	Speed limit (mph)	x_{i16}	55	70	67.67	3.75
L_WIDTH	Lane width (ft)	x_{i17}	12	20	12.06	0.53
NUM_LANE	Number of lanes	x_{i18}	2	8	4.21	0.78
MD_SHOULD	Median width (ft)	x_{i19}	0	60	23.51	29.21

significant at 95% confidence interval were section length, number of lanes, lane width, horizontal curvature, vertical grade, AADT per lane, inside shoulder width, inside rumble strip, and yearly dummy variables for 2005, 2006, 2007, and 2008.

Negative Binomial Regression Model

The negative binomial regression model naturally accounts for the overdispersion, as its variance is greater than the variance of a Poisson distribution. Hence, the model developed in the previous section was reinvestigated using negative binomial assumptions. The maximum likelihood estimates of negative binomial regression model parameters, including the dispersion parameter and goodness-of-fit statistics, are given in Table 3. Both significant and insignificant variables were presented in the table because the primary objective in developing models is to understand the effect of each variable. The sign of the significant variables did not change after removing the insignificant variables from the model.

The dispersion parameter of the estimated negative binomial regression model was 0.5596. Since the dispersion parameter was greater than zero, the response variable was overdispersed. If the deviance value was equal to zero, the model was considered to be a perfect-fit model. Thus, the lowest deviance value was considered to have a better fit. Pearson Chi-Square statistics divided by degree of freedom, and deviance statistics divided by degree of freedom closer to one, indicated a better model fit. Scaled deviance statistics divided by degree of freedom (0.804) were closer to one in the developed negative binomial regression than that of the Poisson regression model (0.771). Hence, the negative binomial model was selected as the better model that can be used to identify the relationship between number of large truck crashes and traffic and geometric-related characteristics on limited access roadways.

Each significant variable in the negative binomial model affected the number of large truck crashes and the magnitude of the coefficient gave the size of the exponential effect of that variable on the number of large truck crashes. The coefficients of continuous variables bearing a plus sign indicate an increase in large truck crashes due to the variable, while a minus sign indicates a decrease in large truck crashes with an increase in the variable. Coefficient of dummy or indicator variables bearing a positive sign indicated that when the dummy or indicator variable switches from 0 to 1 there is an increase in number of crashes. The significant variables in the negative binomial model were section length, number of lanes, horizontal curvature, vertical grade, AADT per lane, large truck percent, inside shoulder width, and annual dummy variables, 2005-2008. The effect on the number of large truck crashes from each of these variables is explained below.

Length of Section: The negative binomial model showed that section length has a positive sign, signifying that for a unit increase in length of a section, crash frequency also increases if all other variables are kept constant. The effect of section length on expected crash frequency showed that shorter sections were less likely to have more large truck crashes than longer sections. This finding was expected and compatible with previous findings on the relationship between length of section and large truck crash frequencies (Miaou 1994, Schneider et al. 2009, Joshua et al. 1990).

Number of Lanes: The variable for number of lanes was significant with a positive coefficient. This means that as the number of lanes increases, opportunities for conflicts related to lane changes also increases, thereby increasing the number of crashes. This was also found by previous researchers (Miaou 1994).

Horizontal Curvature: The horizontal curvature-related variable indicated large truck crashes were less likely on sharp curves. This finding was rather difficult to explain, even though it is compatible with some of the previous findings (Daniel et al. 2002, Milton and Mannering 1998). The variable,

Table 3: Developed Poisson Regression Model and Negative Binomial Model

Variable	Description	Poisson Regression Model		Negative Binomial Model	
		Estimate	P-value	Estimate	P-value
Intercept		-14.260	<0.001	-3.775	*<0.001*
SEC_LEN	Section length (in mile)	*0.1738**	*<0.001*	*0.2231**	*<0.001*
L_WIDTH	Lane width (in ft)	*0.1363**	*0.010*	0.0491	0.348
SPEED	Posted speed limit (in mph)	-0.0068	0.302	0.0067	0.331
NUM_LANE	Number of lanes	*0.0927**	*<0.001*	*0.0652**	*0.013*
HC	Horizontal curvature (in degree per 100ft arc)	*-0.6097**	*<0.001*	*-0.5622**	*<0.001*
VG	Vertical grade	*-0.4348**	*<0.001*	*-0.3916**	*<0.001*
AADT	AADT of the traffic stream per lane	*0.1592**	*<0.001*	*0.2035**	*<0.001*
TRUCK	Large Truck Percent**	-	-	*0.0141*	*<0.001*
R_SHOULD	Right shoulder width in ft	0.0360	0.217	0.0069	0.802
IN_SHOULD	Inside shoulder width in ft	*0.0697**	*<0.001*	*0.0863**	*<0.001*
Y_2005	Dummy variable for year 2005	*0.3706**	*<0.001*	*0.3691**	*<0.001*
Y_2006	Dummy variable for year 2006	*0.2289**	*<0.001*	*0.2266**	*0.002*
Y_2007	Dummy variable for year 2007	*0.1915**	*0.001*	*0.2629**	*<0.001*
Y_2008	Dummy variable for year 2008	*0.1546**	*0.002*	*0.2110**	*0.001*
Y_2009	Dummy variable for year 2009	-0.1209	0.071	-0.0984	0.1404
MD_SHOULD	Median width in ft	-0.0017	0.217	-0.0005	0.698
R_RUMBLE	Dummy variable for right rumble strip	0.0501	0.470	-0.0590	0.447
IN_RUMBLE	Dummy variable for inside rumble strip	*-0.2532**	*0.001*	-0.1071	0.215
Scale		1.2145		-	
Dispersion		-		0.5596	
Goodness-of-Fit Statistics					
Criterion		Value	Value/DF	Value	Value/DF
Deviance		8,256	1.138	5,897	0.813
Pearson Chi-Square		10,701	1.475	7,338	1.011
Scaled Deviance		5,597	0.771	5,830	0.804
Scaled Pearson Chi-Square		7,255	1.000	7,255	1.000
Number of Observations (road sections)		7,273		7,273	

Note: * *Significant values at 95% confidence level*
*** Large truck percent in Poisson regression model was considered as an exposure variable*

horizontal curvature, works in conjunction with the length of the section; hence, the net effect of a horizontal curvature on large truck crash frequencies seems to be inconclusive, as some of the past studies found a positive relationship between large truck crash frequencies and horizontal curvature (Miaou 1994, Mohamedshah et al. 1993, Schneider et al 2009), while others did not.

Vertical Grade: Vertical grade was negatively correlated with large truck crash frequency. One possible explanation was that curves in vertical plane on a limited-access highway consist of minor initial grades and adequate sight distances. The combination of upgrades and downgrades may not be giving a clear relationship between the vertical grades and large truck crash frequencies. However, many previous studies have used the absolute value of vertical grade as an independent variable when modeling the crash frequencies (Miaou 1994, Mohamedshah 1993, Joshua and Garber 1990). The negative relationship between truck crashes and vertical grade was also found by Daniel et al. (2002) when investigating intersection-related crashes.

AADT per Lane: An increase in AADT per lane tended to increase large truck crash frequency. As the number of vehicles through a section increases, exposure to potential crash situations and number of conflicts also increases. This finding was expected, and a relationship was also found in previous studies by Miaou (1994) and Mohamedshan et al. (1993).

Large Truck Percent: Positive coefficient of the large truck percent in the model indicated that as the percentage of large trucks through a section increases, the number of crashes increases. This is consistent with the expectation that the number of truck crashes should increase if there are proportionally more large trucks. Some of the previous research has found that large truck crash frequency decreases with an increase in the percentage of large trucks (Miaou 1994, Milton and Mannering 1998). The explanation in those studies was that the presence of more large trucks reduced vehicle overtaking and lane changing behaviors, which are more crucial for safety. However, if and when the AADT is relatively low, even a few additional trucks on roadways increase the truck percentage (Milton and Mannering 1998). Thus, large truck crashes may decrease in some cases when AADT is low, because of lack of conflicts, not because of an increase in large trucks. Another study has shown that the number of large truck crashes increases with an increase in large truck AADT (Schneider et al. 2009). Hence, large truck percentage works in conjunction with the AADT, making large truck percentage to be another inconclusive variable.

Inside Shoulder Width: Inside shoulder width had a positive correlation with the number of large truck crashes, meaning the number of crashes increases when inside shoulder width increases. A similar relationship has been found by Ivan et al. (1999) when analyzing two-lane rural highways. However, with narrower shoulder widths, drivers have less room to take corrective actions after making an errant maneuver, and drivers are more likely to be involved in fixed-object crashes with the reduced widths. Hence, it was expected to see a decreased number of large truck crashes when shoulder width was increased. So the result was not expected.

Yearly Dummy Variables: The coefficient of yearly variables for 2005, 2006, 2007, and 2008, which represented unmeasured factors, was positive and significant. This means the overall number of large truck crashes increased due to unmeasured factors not included in the model. Similar findings were documented by Miaou (1994).

In this study, absolute values for the variables' vertical grade and horizontal curvature were used because the considered analysis unit includes both directions of travel. A positive gradient value for one direction is a negative for the other direction. One possible way to address this issue is to model the crashes on one direction of travel at a time; however, many previous research studies modeled crashes on both directions considering the absolute values of horizontal curvature (Miaou 1994, Joshua and Garber 1990, Schneider et al. 2009). The reason that the variable on horizontal curvature was inconclusive may also be due to segmentation issues; however, modeling without considering horizontal curves sections did not affect the results on how other variables are affecting the outcome either. Variables such as speed limit, shoulder width, and road width have certain fixed values for each road segment. Hence, these can be defined either as categorical or dummy variables, which might have had some effect on the outcome. Some of the results such as horizontal curvature, vertical grade, and inside shoulder width, from the Poisson regression model and negative binomial model were not as expected. Hence, advanced model formats such as random parameter negative binomial model or Zero-Inflated models may be tested in the next steps, to check whether that might lead to a more robust model.

Based on the developed model, the relationship between large truck crashes and geometric design features, traffic, and other characteristics were identified. The identified effective parameters in large truck crashes can be considered as the criteria for improving highway safety. According to the developed models, it can be concluded that the variables such as number of lanes, AADT, and large truck percent have a specific impact on large truck crashes. Developed models can be used to identify target improvements to limited-access highways to reduce large truck crashes. Also, these can be used to form public policy and highway design criteria. This understanding offers important insight into the relationship between safety and mobility that will improve the quality of decisions made by practicing engineers and planners.

DISCUSSION

In roadway designing, features normally considered are road cross-section elements such as roadway median, utility and landscape areas, drainage channels and side slopes; sight-distance considerations; and horizontal/vertical curvatures as per the design guides. One of the most important factors in design of a limited-access highway facility is design speed. For urban areas, the designer needs to select a reasonable design speed, considering access restrictions and type of access control that can be achieved. Limited-access roadways need to be designed with smooth-flowing horizontal and vertical alignments. Proper combination of horizontal curvature, grades, and median types are expected to provide safety and aesthetics of roadways. The dimensions, weight per axle, and operating characteristics of a vehicle influence design aspects such as width of the lane and curvature. Additionally, consideration of human, traffic, and environmental factors are important in designing roadways as well (Bonneson and Lord 2005).

In recent years, a number of studies have been conducted on geometric design features, safety and operational effect of those designs, and how they influence other activities. The National Cooperative Highway Research Program has reported those findings in Synthesis Report 432 (Brewer 2012). According to the report, large trucks are given important consideration in the geometric design. Some research has given several recommendations for updating existing design guides. Lamm et al. (2002) have developed a process to evaluate the safety of horizontal alignment on two-lane rural roads. This methodology allows designers to predict potential crash risks and safety-related concerns of an alignment, and make changes or develop countermeasures. The occurrence of crashes on two-lane highways is different than on multilane divided highways, but a similar process for evaluating the safety of horizontal alignment on multilane highways may be developed.

Engineers and transportation planners make decisions to add travel lanes on a freeway when they find the capacity of the road needs to increase. According to results of this study, the number of large truck crashes increases when the traffic volume increases. Engineers and planners may believe that decreased traffic is associated with some degree of improved safety. However, results also showed that crashes increase with an increase in the number of lanes. Hence, the introduction of barrier-separated lanes, express lanes, and managed lanes such as toll roadways and dual-dual lanes are effective strategies to offset the increase of conflict opportunities associated with an increase in the number of lanes (Kononov et al. 2008). Dual-dual lanes are managed lanes that have physically separated inner and outer lanes in each direction. The inner lane is reserved for light vehicles, while the outer roadway is open to all vehicles. These lane strategies are a treatment for a specific section of roadway that has a unique set of characteristics such as vertical grades, weaving area, and high percentage of large truck traffic.

The percent increase of large truck traffic is increasing the number of large truck crashes. This is an important matter for all drivers because it affects speed of travel, safety, comfort, and convenience. Hence, many transportation agencies have implemented a variety of countermeasures for large trucks in an attempt to mitigate the effects of increasing large truck traffic. One such example is exclusive truck lanes (Kuhn et al. 2005). California operates an exclusive truck roadway on IH-5 in the Los Angeles area. While other vehicles are allowed to use the roadway, trucks are the primary users. This limited-access road section that includes vertical grades allows slower truck speeds than the free-flow speed of other vehicles, especially in the uphill direction. The Managed Lanes Handbook suggests exclusive barrier-separated truck lanes if truck volumes exceed 30% of the vehicle mix, peak-hour volumes exceed 1,800 vehicles per lane-hour, and off-peak volumes exceed 1,200 vehicles per lane-hour (Kuhn et al. 2005).

The focus of this study was limited to the investigation of the relationship between roadway geometric characteristics and large truck crashes. However, countermeasures for improving safety are not only limited for geometric improvements but also improvements in pavement markings, traffic signs, roadside improvements, lighting, and changing regulations. According to results, the percent increase of truck traffic is increasing the number of truck crashes. To mitigate the effects of increasing truck traffic, exclusive truck lanes can be used. However, just by increasing number of lanes, fewer truck crashes cannot be expected as results showed a positive relationship between number of lanes and truck crash frequency. Table 4 shows a general countermeasure list that could be used to improve the safety of roadways focusing on all possible areas (Washington et al. 2002). For example, if the case of sharper horizontal curves cannot be avoided, countermeasures such as warning signs can be used to provide enough guidance to the driver. Widening and improving clear zones is an alternative countermeasure, which also helps to reduce run-off-road crashes. This may include flattening side slopes, removal of roadside obstacles, and increasing available stopping distance adjacent to the road. As identified in this study, geometric changes such as horizontal alignments decrease large truck crash frequency. Geometric alternations may be considered when other less costly countermeasures are not effective and when the current roadway geometry designs can significantly benefit from improvements. Before implementing countermeasures, the most effective countermeasures and specific conditions for which they are effective need to be identified. The countermeasures related to road geometry and traffic conditions discussed in this paper are related to preventing large truck crashes, but preventing or reducing the number of truck crashes overall improves traffic safety as well. Not all countermeasures can be implemented simultaneously. Also, some countermeasures are less effective when introduced in isolation.

SUMMARY AND CONCLUSIONS

Traffic- and geometric-related data and crash data for limited-access roads were utilized in this study to model or predict large truck crash frequency in Kansas. Data yielded 7,273 homogeneous, limited-access roadway segments which had speed limits of more than 55 mph and lengths of more than 0.25 miles. Poisson and negative binomial regression models were used to estimate the effects of independent variables. According to the coefficients of the developed negative binomial models, large truck crash frequency increased when the length of a section, the number of lanes, AADT per lane, and inside shoulder width increased. Vertical grades were significantly negatively correlated with large truck crash frequency. Also, the overall number of large truck crashes increased due to unmeasured factors that were not in the model.

The results of the negative binomial model may be used for improvement to limited access highways and to prevent or mitigate large truck crashes. Large trucks need to be given important consideration in the geometric design. Revision of existing design guides needs to take into account current dimensions of large trucks and vertical curvature considerations. A process for evaluating the safety of horizontal alignment on multilane highways can be an effective countermeasure. This process allows designers to predict potential crash risks and safety-related concerns of an alignment, and make changes or develop countermeasures. Introduction of exclusive truck lanes,

Table 4: A General Countermeasure List for Improving Roadway Safety

Category	Countermeasure	
Pavement Markings	Add/Upgrade Edgeline	
	Add Raised Pavement Markings (RPMs)	
Traffic Signs	Warning Signs	
	Advisory Speed Signs	
	Chevron Alignment Sign: These Advance Warning Traffic Signs are used to warn drivers of existing or potentially hazardous condition ahead.	
	Post Delineator: These are designed to aid motorists in identifying the alignment of the roadway when utilized in roadside or centerline	
Roadway Improvements	Modify Geometric Alignment	
	Improve Sight Distance without Geometric Realignment	
	Add Turn Lanes/ Barrier-Separated Lanes/Managed Lanes	
	Improve Longitudinal Shoulder	Add/Widen Graded/Stabilized Shoulder
		Pave Existing Graded Shoulder of Suitable Width
		Widen and Pave Existing Paved Shoulder
	Add Rumble Strips	
Roadside Improvements	Install/Upgrade Guardrail	
	Upgrade Guardrail End Treatment/ Add impact Attenuator: These are usually placed in front of fixed structures to reduce the damage to structures, vehicles, and motorists resulting from a motor vehicle crash.	
	Clear Zone Improvements	Widen Clear Zone
		Flatten Side Slope
		Relocate Fixed Object
		Remove Fixed Object
		Convert Object to Breakaway
		Construct Traversable Drainage Structure: Cross-drainage structures with openings (which are used to move water under the roadway) within the clear zone should be made traversable
Lighting	Add Segment Lighting	
	Add Intersection Lighting	
Regulations	Enforcement Speed Limits	

Source: Washington et al. (2002)

barrier-separated lanes, express lanes, and managed lanes such as dual-dual lanes and toll roadways are effective strategies to offset the increase of conflict opportunities associated with an increase in the number of lanes. Warning signs on approaching curves and widening and improving clear zones are countermeasures for decreasing large truck crash involvement. This research provides a step to identifying traffic- and geometric-related factors that contribute to large truck crashes.

Acknowledgements

The authors to would like to express their appreciation to the Mid-American Transportation Center for funding this project and to KDOT staff, especially Tina Cramer, for their help and support in providing data. The appreciation further extends to Steven Buckley, traffic safety engineer at KDOT, who served as the project monitor.

References

Agresti, Alan. *An Introduction to Categorical Data Analysis*. Vol. 423, Wiley-Interscience, 2007.

American Trucking Associations. *Trucking Safety Facts*, 2012. http://northland300.org/Documents/2011%20Trucking%20Safety%20Facts%20ATA.pdf. Accessed on January 13, 2012.

Bonneson, James A. and Dominique Lord. *Role and Application of Accident Modification Factors in the Highway Design Process*. No. FHWA/TX-05/0-4703-2. Texas Transportation Institute, Texas A&M University System, 2005.

Brewer, Marcus A. *Recent Roadway Geometric Design Research for Improved Safety and Operations*. Vol. 432. Transportation Research Board, 2012.

Daniel, Janice, Chuck Tsai, and Steven Chien. "Factors in Truck Crashes on Roadways with Intersections." *Transportation Research Record: Journal of the Transportation Research Board* 1818 (1), (2002): 54-59.

Ivan, John N., Raghubhushan K. Pasupathy, and Paul J. Ossenbruggen. "Differences in Causality Factors for Single and Multi-vehicle Crashes on Two-Lane Roads." *Accident Analysis & Prevention* 31 (6), (1999): 695-704.

Joshua, Sarath C. and Nicholas J. Garber. "Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models." *Transportation Planning and Technology* 15 (1), (1990): 41-58.

Kononov, Jake, Barbara Bailey, and Bryan K. Allery. "Relationships Between Safety and Both Congestion and Number of Lanes on Urban Freeways." *Transportation Research Record: Journal of the Transportation Research Board* 2083 (1), (2008): 26-39.

Kuhn, Beverly T., Ginger Goodin, Andrew Ballard, Marcus Brewer, Robert Brydia, Jodi Carson, and Susan Chrysler. *Managed Lanes Handbook*. No. FHWA/TX-06/0-4160-24. Texas Transportation Institute, Texas A&M University System, 2005.

Lamm, Ruediger, Basil Psarianos, and Salvatore Cafiso. "Safety Evaluation Process for Two-Lane Rural Roads: A 10-year Review." *Transportation Research Record: Journal of the Transportation Research Board* 1796 (1), (2002): 51-59.

Long, J. Scott. *Regression Models for Categorical and Limited Dependent Variables*. Vol. 7. SAGE Publications, Incorporated, 1997.

National Highway Traffic Safety Administration (NHTSA). *Traffic Safety Facts 2009.* No. HS-811 3888. 2009.

Miaou, Shaw-Pin. "The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions." *Accident Analysis & Prevention* 26 (4), (1994): 471-482.

Milton, John and Fred Mannering. "The Relationship Among Highway Geometrics, Traffic-Related Elements and Motor-Vehicle Accident Frequencies." *Transportation* 25 (4), (1998): 395-413.

Mohamedshah, Yusuf M., Jeffery F. Paniati, and Antoine G. Hobeika. "Truck Accident Models for Interstates and Two-Lane Rural Roads." *Transportation Research Record: Journal of The Transportation Research Board* 1407, (1993): 35-41.

Pedan, Alex. "Analysis of Count Data Using the SAS System." *SUGI* 26, (2001): 22-25.

Schneider, William H., Karl Zimmerman, Dan Van Boxel, and Srutha Vavilikolanu. "Bayesian Analysis of the Effect of Horizontal Curvature on Truck Crashes Using Training and Validation Data Sets." *Transportation Research Record: Journal of the Transportation Research Board* 2096 (1), (2009): 41-46.

Stuster, Jack W. *The Unsafe Driving Acts of Motorists in the Vicinity of Large Trucks*, Anacapa Sciences, Inc. Washington D.C., 1999.

Washington, Simon, Karen Dixon, David White, and C. Wu. *Investigation of Fatal Motor Vehicle Crashes on Two-Lane Rural Highways in Georgia*. No. FHWA-GA-02-9905, Georgia Department of Transportation, Office of Materials and Research, 2002.

***Niranga Amarasingha**, Ph.D., is a senior lecturer in transportation engineering at the Sri Lanka Institute of Information Technology. She completed her Ph.D. at Kansas State University and worked there as a post-doctoral research associate before joining the Sri Lanka Institute of Information Technology. Her research focuses on traffic safety and human factors, traffic engineering, modeling of transportation systems, railroad engineering, and multi-criteria decision making.*

***Sunanda Dissanayake**, Ph.D., P.E. is an associate professor attached to the Department of Civil Engineering at Kansas State University. She received her B.Sc. (Eng), M.Eng., and Ph.D. degrees from the University of Moratuwa in Sri Lanka, Asian Institute of Technology in Thailand, and University of South Florida, respectively. Dissanayake has more than two decades of experience in the area of transportation engineering, focusing on both teaching and research in the area of traffic engineering and highway safety. She is a registered professional engineer and actively participates in activities related to professional societies such as Institute of Transportation Engineers, Transportation Research Board, Transportation Research Forum, and American Society of Civil Engineers.*